21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

When Do Gossip Algorithms Converge in Finite Time?

Guodong Shi, Bo Li, Mikael Johansson and Karl Henrik Johansson

***Abstract*— In this paper, we study finite-time convergence of gossip algorithms. We show that there exists a symmetric gossip algorithm that converges in finite time if and only if the number of network nodes is a power of two, while there always exists a globally finite-time convergent gossip algorithm despite the number of nodes if asymmetric gossiping is allowed. For $n = 2^m$ nodes, we prove that a fastest convergence can be reached in mn node updates via symmetric gossiping. On the other hand, for $n = 2^m + r$ nodes with $0 \le r < 2^m$, it requires at least $mn + 2r$ node updates for achieving a finite-time convergence in cooperation with asymmetric interactions.**

***Index Terms*— gossip algorithms, finite-time convergence, computational complexity**

AMS classifications. 68Q25, 68M12, 90B10

I. Introduction

Various gossip algorithms, in which information exchange is always carried out pairwise among the nodes, have been widely used to structure distributed computation, optimization, and signal processing over peer-to-peer, sensor, and social networks [3], [2], [8], [5], [11], [12], [13], [14], [6], [7]. Gossip averaging plays a fundamental role in the study of gossip algorithms due to its simple nature and wide application.

Consider a network with node set $V = \{1, \dots, n\}$. Let the value of node i at time k be $x_i(k) \in \mathbb{R}^1$ for $k \ge 0$. Introduce

$$\mathcal{M} \doteq \Big\{ M_{ij} \doteq I - \frac{(e_i - e_j)(e_i - e_j)^T}{2} : \ i, j = 1, \dots, n \Big\},$$

where $e_m = (0 \dots 0\ 1\ 0 \dots 0)^T$ is the $n \times 1$ unit vector whose m'th component is 1. Denote $x(k) = (x_1(k) \dots x_n(k))^T$. Then a symmetric deterministic gossip algorithm is defined by

$$x(k+1) = P_k x(k), \tag{1}$$

where $\{P_k\}_0^\infty$ satisfies $P_k \in \mathcal{M}$ for all k. Enlarge the set of state-transition matix by [8], [9]

$$\mathcal{M}_* \doteq \Big\{ I - \frac{(e_i - e_j)(e_i - e_j)^T}{2} : \ i, j = 1, \dots, n \Big\}$$
$$\bigcup \Big\{ I - \frac{e_i(e_i - e_j)^T}{2} : \ i, j = 1, \dots, n \Big\}.$$

We call Algorithm (1) an asymmetric gossip algorithm given by $\{P_k\}_0^\infty$ if instead we have $P_k \in \mathcal{M}_*$ for all k.

Algorithm (1) and its variations have been extensively studied in the literature for both randomized and deterministic models. Karp et al. [2] derived a general lower bound for synchronous gossiping; Kempe et al. [3] proposed a randomized gossiping algorithm on complete graphs and determined the order of its convergence rate. Then Boyd et al. [5] established both lower and upper bounds for the convergence time of synchronous and asynchronous randomized gossiping algorithms, and developed algorithms for optimizing parameters to obtain fast consensus. Fagnani and Zampieri discussed asymmetric gossiping in [8] and asymmetric update in random setting was further studied in [9]. Liu et al. [10] presented a comprehensive analysis for the asymptotic convergence rates of deterministic averaging, and recently distributed gossip averaging subject to quantization constraints was studied in [13]. Distributed signal processing and estimation algorithms via gossiping were discussed in [11], [12]. A detailed introduction to gossip algorithms can be found in [6].

In this paper, we study the finite-time convergence of gossip algorithms with its presise definition given as follows.

Definition 1.1: A gossip algorithm in the form of (1) given by $\{P_k\}_0^\infty$ achieves finite-time convergence with respect to initial value $x(0) = x^0 \in \mathbb{R}^n$ if there exists an integer $T(x^0) \ge 0$ such that $x(T) = P_{T-1} \cdots P_0 x(0) \in \mathrm{span}\{\mathbf{1}\}$. Global finite-time convergence is achieved if such $T(x^0)$ exists for every initial value $x^0 \in \mathbb{R}^n$.

We also introduce the definition on the computatonal complexity of finite-time convergent gossiping algorithm.

Definition 1.2: Let Algorithm (1) given by $\{P_k\}_0^\infty$ define a symmetric or asymmetric gossip algorithm. The number of node updates up to T is given by

$$\mathbf{C}_T := \sum_{k=0}^{t-1} \|I_n - P_k\|_1,$$

where $\|\cdot\|_1$ is the matrix norm defined by $\|A\|_1 = \sum_{i=1}^m \sum_{j=1}^n \big|[A]_{ij}\big|$ for any $A \in \mathbb{R}^{m \times n}$ with $\big|\cdot\big|$ denoting the absolute value. The computational complexity of $\{P_k\}_0^\infty$ is indexed by

$$\max_{x^0 \in \mathbb{R}^n} \min_{T \ge 0} \big\{ \mathbf{C}_T : P_{T-1} \cdots P_0 x^0 \in \mathrm{span}\{\mathbf{1}\} \big\}$$

This work has been supported in part by the NSFC of China under Grant 11301518, Knut and Alice Wallenberg Foundation, the Swedish Research Council, and KTH SRA TNG.

G. Shi is with College of Engineering and Computer Science, The Australian National University, ACT 0200 Canberra, Australia. B. Li is with Key Laboratory of Mathematics Mechanization, Academy of Mathematics and Systems Science, and National Center for Mathematics and Interdisciplinary Sciences, Chinese Academy of Sciences, Beijing 100190, China. M. Johansson, and K. H. Johansson are with ACCESS Linnaeus Centre, School of Electrical Engineering, Royal Institute of Technology, Stockholm 10044, Sweden. Email: {guodongs, mikaelj, kallej}@kth.se, libo@amss.ac.cn.

whenever the above equation defines a finite number.

Reaching a consensus in finite-time pushes the convergence rate optimization of gossip algorithms to the limit [5], and by itself it is a basic and fundamental question for distributed gossip computation. We are interested in the following aspects: (i) Is it possible to reach finite-time convergence for gossip algorithms? (ii) What is the essential difference between symmetric and asymmetric gossiping? (iii) Whenever finite-time convergence is possible, what is its computational complexity?

We present clear answers to these questions in the rest of discussions. Section II and Section III will focus on symmetric and asymmetric gossip algorithms, respectively. Some concluding remarks are given in Section IV.

II. Symmetric Gossiping

In this section, we investigate the possiblity and complexity of finite-time convergence for symmetric gossiping algorithms.

We present the following main result on the finite-time convergence of gossip algorithms.

Theorem 2.1: There exists a symmetric gossip algorithm $\{P_k\}_0^\infty$, $P_k \in \mathcal{M}, k \geq 0$, that converges globally in finite time if and only if there exists an integer $m \geq 0$ such that $n = 2^m$. If $n = 2^m$, a fastest symmetric gossip algorithm is reached by mn node updates.

Theorem 2.1 indicates that if the number of nodes n is not some power of two, finding a gossip algorithm which converges globally in finite time is impossible. However, in this case, there still might exist a gossip algorithm which converges in finite time for some initial values, say, half of $\mathbb{R}^n$. The following result further excludes the possibility of the existence of such algorithms by an indeed stronger claim, which shows that the initial values from which there exists a gossip algorithm converging in finite time form a measure zero set.

Theorem 2.2: Suppose there exists no integer $m \geq 0$ such that $n = 2^m$. Then for almost all initial values, it is impossible to find a symmetric gossip algorithm $\{P_k\}_0^\infty$ with $P_k \in \mathcal{M}, k \geq 0$, to reach finite-time convergence.

We give some remarks on randomized algorithms. Most existing works on gossiping algorithms use randomized models [3], [2], [8], [5], [11], [12], [13], [14]. Deterministic gossiping was discussed in [13], [10]. Although we consider deterministic algorithms in this paper, the results can still be easily extended to randomized gossip algorithms.

A. Proof of Theorem 2.1

We prove the necessity, sufficiency, and the fastest convergence statements, respectively.

1) Necessity: Suppose $n = 2^{n_1}n_2$ with $n_1 \geq 0$ and $n_2 \geq 3$ an odd integer. Suppose $P_0, \dots, P_{k_*} \in \mathcal{M}$ with $k_* \geq 0$ gives an algorithm of (1) that converges in finite time globally.

Take $x_1, \dots, x_{2^{n_1}} = 0$ and $x_{2^{n_1}+1}, \dots, x_n = 2^{k_*+1}$. Then there exists $c \in \mathbb{R}$ such that $x_i(k_*+1) = c, i = 1, \dots, n$. On the one hand, because each element in $\mathcal{M}$ is symmetric and therefore doubly stochastic, average is always preserved. Thus, we have

$$c = \frac{2^{k_*+1}2^{n_1}(n_2-1)}{2^{n_1}n_2} = \frac{2^{k_*+1}(n_2-1)}{n_2}.$$

On the other hand, it is not hard to see that c is an integer for the given initial value since pairwise averaging takes place k_*+1 times. Consequently, we have $c = r_2 2^{r_1}$ with $0 \leq r_1 \leq k_*+1$ an integer and $r_2 \geq 1$ an odd integer.

Therefore, we conclude that

$$\frac{2^{k_*+1}(n_2-1)}{n_2} = r_2 2^{r_1},$$

which implies

$$2^{k_*+1-r_1}(n_2-1) = r_2 n_2. \tag{2}$$

This is impossible because the left-hand side of Eq. (2) is an even number while the right-hand side odd. Therefore, (1) cannot achieve global finite-time convergence no matter how $P_0, \dots, P_k, \dots$ are chosen.

2) Sufficiency: We need to construct a gossip algorithm which converges in finite time globally for $n = 2^m$.

We relabel the nodes in a binary system. We use the binary number

$$B_1 \dots B_m, \quad B_s \in \{0,1\}, s = 1, \dots, m$$

to mark node i if $B_1 \dots B_m = i-1$ as a binary number. The gossip algorithm is derived from the following matrix selection process:

S1. *Let $k = 1$.*
S2. *Take 2^{m-1} matrices from $\mathcal{M}$, as the elements in the following set*
$\mathcal{P}_k \doteq \big\{I - \frac{(e_i-e_j)(e_i-e_j)^T}{2}$: in the binary system, the k'th digit of $i-1$ equals 0, and the k'th digit of $j-1$ equals $1\big\}$.
In other words, we take all the node pairs (i,j), where $i-1$ and $j-1$ have identical expressions in the binary system except for the k'th digit. Label the matrices in $\mathcal{P}_k$ as $P^*_{(k-1)2^{m-1}}, \dots, P^*_{k2^{m-1}-1}$ with an arbitrary order.
S3. *Let $k = k+1$ and go to S2 until $k = m$.*

Following this matrix selection process, $P^*_0, \dots, P^*_{m2^{m-1}-1}$ gives a gossip algorithm in the form of (1). It is easy to see that the vector

$$P^*_{s2^{m-1}-1} \cdots P^*_0 x^0, \quad x^0 \in \mathbb{R}^n, \ s = 1, \dots, m$$

has at most 2^{m-s} different elements. Thus, convergence is reached after $m2^{m-1} = (n\log_2 n)/2$ updates. This completes the proof.

3) Complexity: Assume $x_i(0) = a_i$, for $i = 1,\ 2,\ ...,\ 2^m$. Given any gossip algorithm $\{P_k\}_0^\infty$. After multiplication of h matrices the value of every point can be written in the form

$$x_i(h) = \sum_{j=1}^{2^m} \frac{A_{i,h,j}}{2^{B_{i,h,j}}} a_j$$

where $A_{h,j}$ and $B_{h,j}$ are nonnegative integers which depends on $\{P_k\}_0^\infty$ and $\frac{A_{i,h,j}}{2^{B_{i,h,j}}}$ is uniquely determined for all initial values in R^{2^m}.

For any node i, denote $s_{i,h}$ as the times node i has been updated for the initial h matrices.

Claim. $\frac{A_{i,h,i}}{2^{B_{i,h,i}}} \geq \frac{1}{2^{s_{i,h}}}$.

This can be proved by induction on $s_{i,h}$. For $s_{i,h} = 0$, that is to say node i has not been updated for the first h matrices. Then $x_i(h) = a_i$. Thus $\frac{A_{i,h,i}}{2^{B_{i,h,i}}} = 1 = 2^{s_{i,h}}$.

Assume $s_{i,h} = l$, the claim is true. Consider the case $s_{i,h} = l+1$, assume at the multiplication of the h'th matrix, node i is updated for the $(l+1)$-th time. Then, by the induction hypothesis, $\frac{A_{i,h'-1,i}}{2^{B_{i,h'-1,i}}} \geq \frac{1}{2^{s_{i,h'-1}}} = \frac{1}{2^l}$. Assume at matrix $P_{h'-1}$, node i and j are updated, i. e. $P_{h'-1} = I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}$.

$$x_i(h') = \frac{x_i(h'-1) + x_j(h'-1)}{2}.$$

The coefficient of a_i is

$$\left(\frac{A_{i,h'-1,i}}{2^{B_{i,h'-1,i}}} + \frac{A_{j,h'-1,i}}{2^{B_{j,h'-1,i}}}\right)/2$$

which is not less than $\frac{A_{i,h'-1,i}}{2^{B_{i,h'-1,i}}}/2$. That is to say

$$\frac{A_{i,h',i}}{2^{B_{i,h',i}}} \geq \frac{A_{i,h'-1,i}}{2^{B_{i,h'-1,i}}}/2 \geq \frac{1}{2^{s_{i,h'-1}+1}} = \frac{1}{2^{l+1}}$$

For $s_{i,h} = l+1$, node i will not be updated in the rest matrices of the initial h matrices. Thus, $x_i(h) = x_i(h')$. $\frac{A_{i,h,i}}{2^{B_{i,h,i}}} = \frac{A_{i,h',i}}{2^{B_{i,h',i}}} \geq \frac{1}{2^{l+1}} = \frac{1}{2^{s_{i,h}}}$. All the above proved the claim.

For each multiplication, the sum of all nodes is not changed, i. e. for any h

$$\sum_{l=1}^{2^m} x_l(h) = \sum_{l=1}^{2^m} x_l(h+1).$$

Thus, if gossip algorithm $\{P_k\}_0^\infty$ converges at finite matrix P_{T-1},

$$x_1(T) = x_2(T) = \ldots = x_{2^m}(T) = \frac{\sum_{l=1}^{2^m} a_l}{2^m} = \sum_{l=1}^{2^m} \frac{1}{2^m} a_l.$$

According to the claim, $\frac{1}{2^m} = \frac{A_{i,T,i}}{2^{B_{i,T,i}}} \geq \frac{1}{2^{s_{i,T}}}$, for any i. Thus, $s_{i,T} \geq m$. That is to say, when all point converges to the same value, each node must have been updated for at least m times. We know that for each multiplication of matrix only two points are updated. Therefore, T is at least $mn/2$ and thus the least number of node updates equals to mn.

B. *Proof of Theorem 2.2*

The proof is built upon an understanding to the finite-time convergence of the general class of averaging algorithms. In fact, (1) is a special case of distributed averaging algorithms defined by products of stochastic matrices [16], [17], [18]:

$$x(k+1) = W_k x(k), \tag{3}$$

where $W_k \in \mathcal{S} \doteq \big\{W \in \mathbb{R}^{n\times n} : W \text{ is a stochastic matrix}\big\}$. Let $\mathcal{S}_0 \subseteq \mathcal{S}$ be a subset of stochastic matrices. We define $\mathscr{X}_{\mathcal{S}_0} \doteq \Big\{x \in \mathbb{R}^n : \exists W_0, \ldots, W_s \in \mathcal{S}_0, s \geq 0 \text{ s.t. } W_s \cdots W_0 x \in \operatorname{span}\{\mathbf{1}\}\Big\}$.

Let $\mathbf{M}(\cdot)$ represent the standard Lebesgue measure on $\mathbb{R}^n$. We have the following result for the finite-time convergence of general averaging algorithms.

Proposition 2.1: Suppose $\mathcal{S}_0$ is a set with at most countable elements. Then either $\mathscr{X}_{\mathcal{S}_0} = \mathbb{R}^n$ or $\mathbf{M}(\mathscr{X}_{\mathcal{S}_0}) = 0$. In fact, if $\mathscr{X}_{\mathcal{S}_0} \neq \mathbb{R}^n$, then $\mathscr{X}_{\mathcal{S}_0}$ is a union of at most countably many linear spaces whose dimensions are no larger than $n-1$.

Remark 2.1: Note that in the definition of $\mathscr{X}_{\mathcal{S}_0}$, different initial values can correspond to different averaging algorithms. Even if $\mathcal{S}_0$ is finite, there will still be uncountably many different averaging algorithms in the form of (3) as long as $\mathcal{S}_0$ contains at least two elements. Therefore, the proof of Proposotion 2.1 requres a careful strcture characterization to $\mathscr{X}_{\mathcal{S}_0}$.

Proof of Proposition 2.1. Define a function $\delta(M)$ of a matrix $M = [m_{ij}] \in \mathbb{R}^{n\times n}$ by (cf. [15])

$$\delta(M) \doteq \max_j \max_{\alpha,\beta} |m_{\alpha j} - m_{\beta j}|. \tag{4}$$

Given an averaging algorithm (3) defined by $\{W_k\}_0^\infty$ with $W_k \in \mathcal{S}_0, k \geq 0$. Suppose there exists an initial value $x^0 \in \mathbb{R}^n$ for which $\{W_k\}_0^\infty$ fails to achieve finite-time convergence. Then obviously $\delta(W_s \cdots W_0) > 0$ for all $s \geq 0$.

Claim. $\operatorname{rank}(W_s \cdots W_0) \geq 2,\ s \geq 0$.

Let $W_s \cdots W_0 = (\omega_1 \ldots \omega_n)^T$ with $\omega_i \in \mathbb{R}^n$. Since $\delta(W_s \cdots W_0) > 0$, there must be two rows in $W_s \cdots W_0$ that are not equal. Say, $\omega_1 \neq \omega_2$. Note that $W_s \cdots W_0$ is a stochastic matrix because any product of stochastic matrices is still a stochastic matrix. Thus, $\omega_i \neq 0$ for all $i = 1, \ldots, n$. On the other hand, if $\omega_1 = c\omega_2$ for some scalar c, we have $1 = \omega_1^T \mathbf{1} = c\omega_2^T \mathbf{1} = c$, which is impossible because $\omega_1 \neq \omega_2$. Therefore, we conclude that $\operatorname{rank}(W_s \cdots W_0) \geq \operatorname{rank}(\operatorname{span}\{\omega_1, \omega_2\}) \geq 2$. The claim holds.

Suppose there exists some $y \in \mathbb{R}^n$ such that $y \notin \mathscr{X}_{\mathcal{S}_0}$. We see from the claim that the dimension of $\ker(W_s \cdots W_0)$ is at most $n-2$ for all $s \geq 0$ and $W_0, \ldots, W_s \in \mathcal{S}_0$.

Now for $s = 0, 1, \ldots$, introduce $\Theta_s \doteq \big\{x \in \mathbb{R}^n : \exists W_0, \ldots, W_s \in \mathcal{S}_0, \text{ s.t. } W_s \cdots W_0 x \in \operatorname{span}\{\mathbf{1}\}\big\}$. Then Θ_s indicates the initial values from which convergence is reached in $s+1$ steps. For any fixed $W_0, \ldots, W_s \in \mathcal{S}_0$, we define

$$\Upsilon_{W_s \ldots W_0} \doteq \big\{z \in \mathbb{R}^n : \quad W_s \cdots W_0 z \in \operatorname{span}\{\mathbf{1}\}\big\}.$$

Clearly $\Upsilon_{W_s \ldots W_0}$ is a linear space. It is straightforward to see that $\Theta_s = \bigcup_{W_s \ldots W_0 \in \mathcal{S}_0} \Upsilon_{W_s \ldots W_0}$, and therefore

$$\mathscr{X}_{\mathcal{S}_0} = \bigcup_{s=0}^{\infty} \Theta_s = \bigcup_{s=0}^{\infty} \bigcup_{W_s, \ldots, W_0 \in \mathcal{S}_0} \Upsilon_{W_s \ldots W_0}.$$

Noticing that $z \in \Upsilon_{W_s \dots W_0}$ implies $\big(z - W_s \cdots W_0 z\big) \in \ker(W_s \cdots W_0)$, we define a linear mapping

$$\begin{aligned} f: \quad & \Upsilon_{W_s \dots W_0} \longmapsto \ker(W_s \cdots W_0) \times \text{span}\{\mathbf{1}\} \\ \text{s.t.} \quad & f(z) = \big(z - W_s \cdots W_0 z, W_s \cdots W_0 z\big) \end{aligned} \tag{5}$$

Suppose $z_1, z_2 \in \Upsilon_{W_s \dots W_0}$ with $z_1 \neq z_2$. It is straightforward to see that either $W_s \cdots W_0 z_1 = W_s \cdots W_0 z_2$ or $W_s \cdots W_0 z_1 \neq W_s \cdots W_0 z_2$ implies $f(z_1) \neq f(z_2)$. Hence, f is injective. Therefore, noting that $\ker(W_s \cdots W_0)$ is a linear space with dimension at most $n-2$, we have $\dim(\Upsilon_{W_s \dots W_0}) \leq n-1$, and thus $\mathbf{M}(\Upsilon_{W_s \dots W_0}) = 0$. Consequently, we conclude that

$$\begin{aligned} \mathbf{M}(\Theta_s) &= \mathbf{M}\Big(\bigcup_{W_0, \dots, W_s \in \mathcal{S}_0} \Upsilon_{W_s \dots W_0}\Big) \\ &\leq \sum_{W_0, \dots, W_s \in \mathcal{S}_0} \mathbf{M}\big(\Upsilon_{W_s \dots W_0}\big) \\ &= 0 \end{aligned}$$

because any finite power set $\mathcal{S}_0 \times \cdots \times \mathcal{S}_0$ is still a countable set as long as $\mathcal{S}_0$ is countable. This immediately leads to

$$\mathbf{M}(\mathscr{X}_{\mathcal{S}_0}) = \mathbf{M}\Big(\bigcup_{s=0}^{\infty} \Theta_s\Big) \leq \sum_{s=0}^{\infty} \mathbf{M}(\Theta_s) = 0.$$

Additionally, since every Θ_s is a union of at most countably many linear spaces, each of dimension no more than $n-1$, $\mathscr{X}_{\mathcal{S}_0}$ is also a union of countably many linear spaces with dimension no more than $n-1$. The desired conclusion thus follows. □

Noticing that $\mathcal{M}$ is a finite set and utilizing Proposition 2.1, Theorem 2.2 follows immediately.

C. Discussion: How Many Algorithms can be Found?

In this subsection, we make some further discussions on essentially how many *different* finite-time convergent algorithms via symmetric gossiping exist. We present the following result indicating that when $n = 4$, the desired algorithm is indeed unique. Recall that $M_{ij} \doteq I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}$. Since the proof of this proposition is rather technical, we refer [19] for a complete proof.

Proposition 2.2: Let $n = 4$. Suppose $P_{T-1} \cdots P_0 = \mathbf{1}\mathbf{1}^T/4$ with $P_{T-2} \cdots P_0 \neq \mathbf{1}\mathbf{1}^T/4$. Then there are under certain permutation of index we always have $P_{T-1} = M_{12}$, $P_{T-2} = M_{34}$, $P_{T-3} = M_{13}$ and $P_{T_\alpha} = M_{24}$ for some $0 \leq T_\alpha < T-3$.

III. Asymmetric Gossiping

In this section, we investigate asymmetric gossiping. It turns out that finite-time convergence is always possible despite the number of nodes as long as asymmetric gossiping is allowed. The following conclusion holds.

Theorem 3.1: There always exists a deterministic gossip algorithm $\{P_k\}_0^\infty$, $P_k \in \mathcal{M}_*, k \geq 0$, which converges globally in finite time. In fact, for $n = 2^m + r$ with $0 \leq r < 2^m$, a fastest asymmetric gossiping algorithms that converges globally in finite time requires $mn + 2r$ node updates.

A. Complexity

In this subsection, we first establish the least number of node updates for finite-time convergence via asymmetric gossiping. For any n, n can be written as $n = 2^m + r$, where m and r are integers and $0 \leq r < 2^m$. The complexity proof relies on the following lemma, whose proof can be found in [19].

Lemma 3.1: Let $n = 2^m + r$ with $0 \leq r < 2^m$. F is a subset of $\mathbb{R}^n$ such that $f = (f_1, \ \dots, \ f_n) \in F$ if and only if

$$1 = \sum_{i=1}^{n} f_i$$

and f_is have the form $\frac{b_i}{2^{c_i}}$ where b_is are positive odd integers and c_is are nonnegative integers, for $i = 1, \ \dots, \ n$. As b_i and c_i are uniquely determined by f, we denote them by $b_i(f)$ and $c_i(f)$ respectively. For each f_i, there exist a smallest positive integer $n_i(f)$ such that $f_i \geq \frac{1}{2^{n_i(f)}}$. Define $\hat{n}(f) = \sum_{i=1}^{n} n_i(f)$. Then,

$$\min_{f \in F} \hat{n}(f) = mn + 2r.$$

B. Existence

We now construct an algorithm that when node states converge to the same value, only $nm + 2r$ node updates have been taken.

Again, we relabel the nodes in a binary system. We use the binary number

$$B_1 \dots B_{m+1}, \quad B_s \in \{0, 1\}, s = 1, \dots, m+1$$

to mark node i if $B_1 \dots B_{m+1} = i - 1$ as a binary number. The asymmetric gossip algorithm is derived from the following matrix selection process:

S1. *Take r matrices from $\mathcal{M}_*$, as the elements in the following set $\mathcal{P}_1 \doteq \Big\{I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}$: $i-1$ and $j-1$ have identical expressions in the binary system except for the 1'st digit.$\Big\}$. Label the matrices in $\mathcal{P}_1$ as $P_0^*, \dots, P_{r-1}^*$ with an arbitrary order.*

S2. *Let $k = 2$.*

S3. *Take r matrices from $\mathcal{M}_*$, as the elements in the following set $\mathcal{P}_{(1,k)} \doteq \big\{I - \frac{e_i(e_i - e_j)^T}{2}$: in the binary system, the 1'th digit of $i-1$ equals 1, and the 1'th digit of $j-1$ equals 0, $i-1$ and $j-1$ have identical expressions in the binary system except for the 1'st and k'th digits. $\big\}$. Label the matrices in $\mathcal{P}_{(1,k)}$ as $P^*_{(k-1)r+(k-2)2^{m-1}}, \dots, P^*_{(k-1)r+(k-2)2^{m-1}+r-1}$ with an arbitrary order.*

S4. *Take 2^{m-1} matrices from $\mathcal{M}_*$, as the elements in the following set $\mathcal{P}_{(2,k)} \doteq \big\{I - \frac{(e_i - e_j)(e_i - e_j)^T}{2}$: $i-1$ and $j-1$ have identical expressions in the binary system except for the k'th digit, and the 1'st digits of $i-1$ and $j-1$ are both $0\big\}$. Label the matrices in $\mathcal{P}_{(2,k)}$ as $P^*_{kr+(k-2)2^{m-1}}, \dots, P^*_{kr+(k-2)2^{m-1}+2^{m-1}-1}$ with an arbitrary order.*

S5. *Let $k = k+1$ and go to S2 until $k = m+1$.*

Following this matrix selection process, $P_0^*, \dots, P^*_{m2^{m-1}+(m+1)r-1}$ gives a asymmetrical gossip algorithm in the form of (1). It is easy to see that the vector

$$P^*_{sr+(s-1)2^{m-1}-1} \cdots P_1^* P_0^* x^0, \quad x^0 \in \mathbb{R}^n, \ s = 1, \dots, m+1$$

has at most 2^{m+1-s} different elements. Note that the matrix selected in S1 and S4 contribute two updated values, and the matrix selected in S3 contribute one updated value. Thus, convergence is reached after $r * 2 + (2^{m-1} * 2 + r) * m = mn + 2r$ value updates. This completes the proof. □

C. Discussion: Fastest Algorithm in Term of Matrices

Here, we choose the number of node value updates as the efficiency of the asymmetrical gossip algorithm instead of the number of matrices selected. It is still unknown the least number of matrices needed to converge. In fact, the algorithm in the above proof does not have the least number of matrices. For example, for $n = 6$, the algorithm selects 10 matrices. However, the following algorithm selects only 9 matrices.

We give the new algorithm by recursion. Denote $\mathscr{A}_n$ as the algorithm defined on n nodes. For $n = 2$, the algorithm is just update each value to their average, which contains 2 updates of values. For $n = 3$, first updates the value of node 1 and 2 by their average value. Then, updates node 1 by average value of node 1 and 3. Finally, updates node 2 and 3 by their average value. This algorithm contains 5 updates of values.

For $n = 2^m + r$, and if $r = 2r_1$, define $\mathscr{A}_n$ as follows. First, take algorithm $\mathscr{A}_{n/2}$ on nodes $1,\ 2,\ \dots,\ n/2$. Then all these $n/2$ nodes converges to the same value. The number of updates in this process is $\frac{n}{2}(m-1) + 2r_1$. Second, take algorithm $\mathscr{A}_{n/2}$ on nodes $n/2+1,\ n/2+2,\ \dots,\ n$. Thus, all these remaining nodes converges. The number of updates in this process is also $\frac{n}{2}(m-1) + 2r_1$. Third, for $i = 1,\ 2,\ \dots,\ n/2$, update the values of nodes i and $i+n/2$ by taking their average. The number of updates of this process is n. Therefore, the whole number of updates for this algorithm is $\frac{n}{2}(m-1) + 2r_1 + \frac{n}{2}(m-1) + 2r_1 + n = nm + 2r$.

For $n = 2^m + r$, and if $r = 2r_2 + 1$, define $\mathscr{A}_n$ as follows. First, take algorithm $\mathscr{A}_{(n-1)/2}$ on nodes $1,\ 2,\ \dots,\ (n-1)/2$. Then all these $(n-1)/2$ nodes converges. The number of updates in this process is $\frac{n-1}{2}(m-1) + 2r_2$. Second, take algorithm $\mathscr{A}_{(n+1)/2}$ on nodes $(n-1)/2+1,\ n/2+2,\ \dots,\ n$. Thus, all these remaining nodes converges. The number of updates in this process is $\frac{n+1}{2}(m-1) + 2(r_2+1)$. Third, update the value of node n by taking the average of node 1 and node n. Fourth, for $i = 1,\ 2,\ \dots,\ (n-1)/2$, update the values of nodes i and $i + (n-1)/2$ by taking their average. The number of updates of this step is $(n-1)$. Therefore, the whole number of updates for this algorithm is $\frac{n-1}{2}(m-1) + 2r_2 + \frac{n+1}{2}(m-1) + 2(r_2+1) + 1 + (n-1) = nm + 2r$.

We provide a conjecture that the algorithm given here has the least number of matrices needed to converge.

IV. Conclusions

We have answered the question on when gossip algorithms admit a convergence in finite time. We showed that there exists a symmetric gossip algorithm that converges in finite time if and only if the number of network nodes is a power of two, while there always exists a globally finite-time convergent gossip algorithm despite the number of nodes if asymmetric gossiping is allowed. In both cases we have constructed desired algorithms explicitly, and we proved that the given algorithms indeed reach fastest convergence. More challenges lie in how to present a precise description on how the graph structure influences the existence and complexity of finite-time convergence via gossiping.

References

[1] G. Latouche, V. Ramaswami. *Introduction to Matrix Analytic Methods in Stochastic Modeling.* 1st edition, ASA SIAM, 1999.

[2] R. Karp, C. Schindelhauer, S. Shenker, and B. Vöcking, "Randomized rumor spreading," in *Proc. Symp. Foundations of Computer Science*, pp. 564-574, 2000.

[3] D. Kempe, A. Dobra, and J. Gehrke, "Gossip-based computation of aggregate information," in *Proc. Conf. Foundations of Computer Science*, pp. 482-491, 2003.

[4] S. Boyd, P. Diaconis and L. Xiao, "Fastest mixing Markov chain on a graph," *SIAM Review*, Vol. 46, No. 4, pp. 667-689, 2004.

[5] S. Boyd, A. Ghosh, B. Prabhakar and D. Shah, "Randomized gossip algorithms," *IEEE Trans. Information Theory*, vol. 52, no. 6, pp. 2508-2530, 2006.

[6] D. Shah, "Gossip Algorithms," *Foundations and Trends in Networking*, vol. 3, no. 1, pp. 1-125, 2008.

[7] D. Mosk-Aoyama and D. Shah, "Fast distributed algorithms for computing separable functions," *IEEE Transactions on Information Theory*, vol.55, no.7, pp. 2997-3007, 2008

[8] F. Fagnani and S. Zampieri, "Asymmetric randomized gossip algorithms for consensus," *IFAC World Congress*, Seoul, pp. 9051-9056, 2008.

[9] F Iutzeler, P Ciblat, W Hachem, "Analysis of Sum-Weight-like algorithms for averaging in Wireless Sensor Networks," *IEEE Transactions on Signal Processing* 61(11), pp. 2802-2814, 2012.

[10] J. Liu, S. Mou, A. S. Morse, B. D. O. Anderson, and C. Yu, "Deterministic gossiping," *Proceedings of IEEE*, vol. 99, no. 9, pp. 1505-1524, 2011.

[11] A. G. Dimakis, S. Kar, J. M. F. Moura, M. G. Rabbat, and A. Scaglione, "Gossip algorithms for distributed signal processing," *Proceedings of IEEE*, vol. 98, no. 11, pp. 1847-1864, 2010.

[12] S. Kar and J. M. F. Moura, "Convergence rate analysis of distributed gossip (linear parameter) estimation: fundamental limits and trade-offs," *IEEE Journal of Selected Topics in Signal Processing*, vol. 5, no. 4, pp.674-690, 2011.

[13] J. Lavaei and R.M. Murray, "Quantized consensus by means of gossip algorithm," *IEEE Trans. Autom. Control*, vol. 57, no.1, pp. 19-32, 2012.

[14] F. Bénézit, A. G. Dimakis, P. Thiran, and M. Vetterli, "Order-optimal consensus through randomized path averaging," *IEEE Transactions on Information Theory*, vol. 56, no. 10, pp. 5150-5167, 2010.

[15] J. Hajnal, "Weak ergodicity in non-homogeneous markov chains," *Proc. Cambridge Philos. Soc.*, no. 54, pp. 233-246, 1958.

[16] J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Trans. Autom. Control*, vol. 31, pp. 803-812, 1986.

[17] A. Jadbabaie, J. Lin, and A. S. Morse, "Coordination of groups of mobile autonomous agents using nearest neighbor rules," *IEEE Trans. Autom.Control*, vol. 48, no. 6, pp. 988-1001, 2003.

[18] V. Blondel, J. M. Hendrickx, A. Olshevsky and J. Tsitsiklis, "Convergence in multiagent coordination, consensus, and flocking," *IEEE Conf. Decision and Control*, pp. 2996-3000, 2005.

[19] G. Shi, B. Li, M. Johansson and K. H. Johansson, "When do gossip algorithms converge in finite time?" arXiv:1206.0992.